


Article

Extending Normality: A Case of Unit Distribution Generated from the Moments of the Standard Normal Distribution

Miguel S. Concha-Aracena¹, Leonardo Barrios-Blanco¹, David Elal-Olivero¹, Paulo Henrique Ferreira da Silva^{2,3} and Diego Carvalho do Nascimento^{1,*} 

¹ Departamento de Matemática, Facultad de Ingeniería, Universidad de Atacama, Copiapó 1530000, Chile

² Department of Statistics, Federal University of Bahia, Salvador 40170110, Brazil

³ Centro de Pesquisa em Matemática Aplicada à Indústria (CeMEAI), University of São Paulo, São Carlos 13566590, Brazil

* Correspondence: diego.nascimento@uda.cl

Abstract: This paper presents an important theorem, which shows that, heading from the moments of the standard normal distribution, one can generate density functions originating a family of models. Additionally, we discussed that different random variable domains are achieved with transformations. For instance, we adopted the moment of order two, from the proposed theorem, and transformed it, which enabled us to exemplify this class as a unit distribution. We named it as Alpha-Unit (AU) distribution, which contains a single positive parameter α ($AU(\alpha) \in [0, 1]$). We presented its properties and demonstrated two estimation methods for the α parameter, the maximum likelihood estimator (MLE) and uniformly minimum-variance unbiased estimator (UMVUE) methods. In order to analyze the statistical consistency of the estimators, a Monte Carlo simulation study was carried out, in which the robustness was demonstrated. As a real-world application, we adopted two sets of unit data, the first regarding the dynamics of Chilean inflation in the post-military period, and the other one regarding the daily maximum relative humidity of the air in the Atacama Desert. In both cases presented, the AU model is competitive, whenever the data present a range greater than 0.4 and extremely heavy asymmetric tail. We compared our model with other commonly used unit models, such as the beta, Kumaraswamy, logit-normal, simplex, unit-half-normal, and unit-Lindley distributions.

Keywords: asymmetry accommodation; rates and proportions; single-parameter distribution; unit distribution; water monitoring



Citation: Concha-Aracena, M.S.; Barrios-Blanco, L.; Elal-Olivero, D.; Ferreira da Silva, P.H.; Nascimento, D.C.d. Extending Normality: A Case of Unit Distribution Generated from the Moments of the Standard Normal Distribution. *Axioms* **2022**, *11*, 666. <https://doi.org/10.3390/axioms11120666>

Academic Editors: Nuno Bastos, Touria Karite and Amir Khan

Received: 20 October 2022

Accepted: 17 November 2022

Published: 24 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Statistical methodology plays an important role in quantitative methods, given the hypothesis testing and inferential procedures. Nonetheless, the comparison across features is given based on a generated function estimated from the data information. Most often, mild suppositions are assumed, which compromises the generalization of the results.

Under the perspective of statistical generalization (inferential method), some challenges are found for bounded distribution estimation. For instance, the confidence interval, which is often adopted from the maximum likelihood estimation approach and asymptotic supposition, is also assumed. Specially, interval estimation can be seen as the parameter space domain.

One exemplification is the case in which bounded information data are observed and, nonetheless, normality is commonly assumed to be true. This is the case of proportion/rate data, which are double bounded in the lower limit equal to zero and upper limit equal to one. Relative humidity is an example of this scenario in which every decision-making should be $\in [0, 1]$ [1,2], or rates commonly used in the fields of finance, economics and demography, to number a few.

In the case of rates and proportions processes, as well as other processes whose variables of interest assume values in the range $(0, 1)$, there is a well-represented class of models, the unit distributions family, which deals with this type of double-bounded data. Among the many existing unit distributions, it is noteworthy mentioning the power distribution, beta distribution [3], Kumaraswamy distribution [4], unit-logistic distribution [5], simplex distribution [6], unit-Weibull distribution [7,8], unit-Lindley distribution [9], unit-half-normal distribution [10], unit log-log distribution [11], modified Kumaraswamy and reflected modified Kumaraswamy distributions [12], unit-Teissier distribution [13], unit extended Weibull families of distributions [14], lognormal distribution [15], unit folded normal distribution [16], Marshall-Olkin reduced Kies distribution [17], and unit-Chen distribution [18].

Despite the applicability of the unit distributions in double-bounded variables, another important fact is that the interval estimation for the parameter may also be limited in a domain (like positive real number). In the face of it, we also presented an inferential alternative through the delta method.

This study starts with a presentation of an important theorem that changes from a modification of the standard normal distribution into a class of density functions that can be seen as a unit. Then, as an exemplification, a second moment case was chosen to illustrate the usefulness of this class of probabilistic models. This class of distributions shows to be competitive for high-frequency data with range greater than 0.4, important to real-world applications, whereas a classical unit distribution fails [19]. Additionally, two different data sets were selected to illustrate the adjustment of the proposed model. The first one is related to the Chilean inflation (ultimate post-military era), and the second one comes from the driest area of the planet (excluding the north and south poles).

This paper is structured in four parts. Section 2 presents the proposed one-parameter unit distribution. In Section 3, the inferences for the distribution parameter adopting the uniformly minimum-variance unbiased estimator (UMVUE) and maximum likelihood estimator (MLE) as point estimators, as well as interval estimations, are discussed. A simulation study is also presented in this section. In Section 4, two real data sets are used to illustrate the proposed methodology, one from the Chilean inflation in the post-military period, and other one from the relative humidity water monitoring in the Atacama Desert. Finally, Section 5 lists the conclusions of this study. Nevertheless, before moving on into the described structure, a wide class of models that can be generated in many different random variable supports is presented. Therefore, a theorem is elicited and, as a special case, the whole paper will consider an order two for exemplification of this powerful class of distributions.

Motivation

The normal (or Gaussian) distribution is very important to the history of statistics, and numerous modifications to this distribution have been proposed in the literature [20,21]. An interesting fact related to the normal distribution is that its even moments can be used to generate new distributions, which is the case presented below, through a definition and a result embodied in a theorem that accounts for the characterization of these new distributions.

Definition 1. A random variable B is said to be distributed according to a Bimodal Normal (BN) distribution of order k , that is, $B \sim BN(k)$ (discussed in [22]), if its probability density function (PDF) is given by

$$f(b | k) = \frac{1}{c} b^{2k} \phi(b), \quad b \in \mathbb{R},$$

in which $\phi(\cdot)$ is the PDF of the standard normal distribution, $c = \prod_{j=1}^k (2j - 1)$ and $k = \{1, 2, 3, \dots\}$.

This class of distributions is always bimodal, which means that the observed modes move away from each other when the order k increases (as depicted by Figure 1).

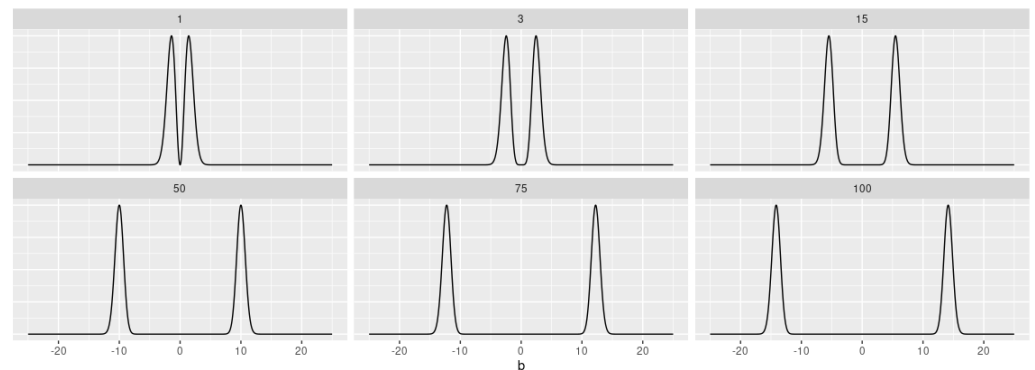


Figure 1. Density function of the BN distribution by varying the parameter k (displayed at the top of each chart).

It is noteworthy mentioning that transformations derived from the $BN(k)$ distribution may lead to other domains of interest, e.g., the unit domain. For example, let $B \sim BN(k)$, then a scale parameter α , the transformation $\alpha|B| \in \mathbb{R}^+$, and then the transformation $e^{-\alpha|B|} \in [0, 1]$. Thus, the stochastic characterization of a $BN(k)$ distribution can be obtained according to the following theorem.

Theorem 1. Let W_1 and W_2 be independent random variables, in which W_1 is such that $\mathbb{P}(W_1 = 1) = \mathbb{P}(W_1 = -1) = 1/2$ and $W_2 \sim \chi^2_{2k+1}$. Then,

$$W_1 \sqrt{W_2} \sim BN(k). \tag{1}$$

So, this theorem is mainly motivated by the result that shows that if $X \sim BN(k)$, then $X^2 \sim \chi^2_{2k+1}$. The entire demonstration is presented in Appendix A.

2. The Model

In this section, a new unit distribution, named Alpha-Unit, which presents a single parameter, α , is discussed. Its stochastic representations (probability density and cumulative distribution functions), moments (including mean and variance), moment-generating function, and how to generate pseudo-random numbers from it will be presented. Moreover, a proposal of statistical control chart for unit data based on the Alpha-Unit distribution will also be shown.

The Alpha-Unit density is originated from the general theorem (Theorem 1), by considering $k = 1$. Moreover, it represents the second moment of the standard normal distribution and, later, transformed its domain. However, as k increases, the concentration of the distribution intensifies and other densities could be obtained.

Properties and Characterization

Definition 2. (Alpha-Unit distribution). A random variable X follows an Alpha-Unit (AU) distribution with parameter $\alpha > 0$, that is, $X \sim AU(\alpha)$, if its PDF is given by

$$f_X(x | \alpha) = \frac{2}{x\alpha} \left(\frac{\ln(x)}{\alpha} \right)^2 \phi \left(\frac{\ln(x)}{\alpha} \right), \quad 0 < x \leq 1. \tag{2}$$

Remark 1. If $X \sim AU(\alpha)$, then its PDF is unimodal.

Proof. The maxima of the AU distribution are studied, to which the criterion of the first derivative is first considered:

$$\frac{df_X(x | \alpha)}{dx} = \frac{2}{x\alpha^2} \frac{\ln(x)}{\alpha} \phi\left(\frac{\ln(x)}{\alpha}\right) \left[\frac{2}{x} - \frac{\ln(x)}{x} - \frac{[\ln(x)]^2}{\alpha} \frac{1}{x\alpha} \right] = 0.$$

By solving algebraically for x , we obtain:

$$x = \begin{cases} e^{\left(\frac{\alpha^2 + \sqrt{\alpha^4 + 8\alpha^2}}{2}\right)} & \text{(i)} \\ e^{\left(\frac{\alpha^2 - \sqrt{\alpha^4 + 8\alpha^2}}{2}\right)} & \text{(ii)} \end{cases}.$$

By working algebraically, it can be seen that this is only true for (ii), and is a global maximum, given that the solution is in between 0 and 1. Therefore, the AU distribution is unimodal. □

Proposition 1. If $X \sim AU(\alpha)$, then its r -th order moment is given by

$$\mathbb{E}[X^r] = 2e^{\left(\frac{r^2\alpha^2}{2}\right)} \left[(1 + r^2\alpha^2)(1 - \Phi(r\alpha)) - r\alpha\phi(r\alpha) \right],$$

in which $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard normal distribution.

Proof. From the definition of the r -th order moment, we have:

$$\mathbb{E}[X^r] = \int_0^1 x^r f_X(x | \alpha) dx = \int_0^1 x^r \frac{2}{x\alpha} \left(\frac{\ln(x)}{\alpha}\right)^2 \phi\left(\frac{\ln(x)}{\alpha}\right) dx. \tag{3}$$

By changing the variables:

$$\begin{cases} u = \frac{1}{\alpha} \ln(x) & \Rightarrow e^{u\alpha} = x \\ du = \frac{1}{\alpha x} dx & \Rightarrow \alpha e^{u\alpha} du = dx \end{cases},$$

then substituting into Equation (3) and developing algebraically, we obtain:

$$\mathbb{E}[X^r] = 2e^{\frac{\alpha^2 r^2}{2}} \int_{-\infty}^0 u^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(u-\alpha r)^2}{2}} du.$$

Then, by making another change of variables: $h = u - \alpha r$, $dh = du$; and replacing these expressions in the previous equation, we have:

$$\begin{aligned} \mathbb{E}[X^r] &= 2e^{\frac{\alpha^2 r^2}{2}} \int_{-\infty}^{-\alpha r} (h + \alpha r)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{h^2}{2}} dh \\ &= 2e^{\frac{\alpha^2 r^2}{2}} \int_{-\infty}^{-\alpha r} (h^2 + 2h\alpha r + \alpha^2 r^2) \phi(h) dh \\ &= 2e^{\frac{\alpha^2 r^2}{2}} \left(\int_{-\infty}^{-\alpha r} h^2 \phi(h) dh + 2\alpha r \int_{-\infty}^{-\alpha r} h \phi(h) dh + \alpha^2 r^2 \int_{-\infty}^{-\alpha r} \phi(h) dh \right). \end{aligned}$$

By solving the integrals, we get to:

$$\mathbb{E}[X^r] = 2e^{\frac{\alpha^2 r^2}{2}} \left[\alpha r \phi(\alpha r) + (1 - \Phi(\alpha r)) - 2\alpha r \phi(\alpha r) + \alpha^2 r^2 (1 - \Phi(\alpha r)) \right].$$

Then, by solving algebraically, we go down to the expression of Proposition 1. □

Out of Proposition 1, we obtain the mean and variance of the $AU(\alpha)$ model as it follows:

$$\begin{aligned} \mathbb{E}[X] &= 2e^{\frac{\alpha^2}{2}} \left[(1 + \alpha^2)(1 - \Phi(\alpha)) - \alpha\phi(\alpha) \right], \\ \text{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= 2e^{2\alpha^2} \left[(1 + 4\alpha^2)(1 - \Phi(2\alpha)) - 2\alpha\phi(2\alpha) \right] - 4e^{\alpha^2} \left[(1 + \alpha^2)(1 - \Phi(\alpha)) - \alpha\phi(\alpha) \right]^2. \end{aligned}$$

Remark 2. As an illustration, Figure 2 displays the generated asymmetry and kurtosis based on the chosen α parameter of the AU distribution.

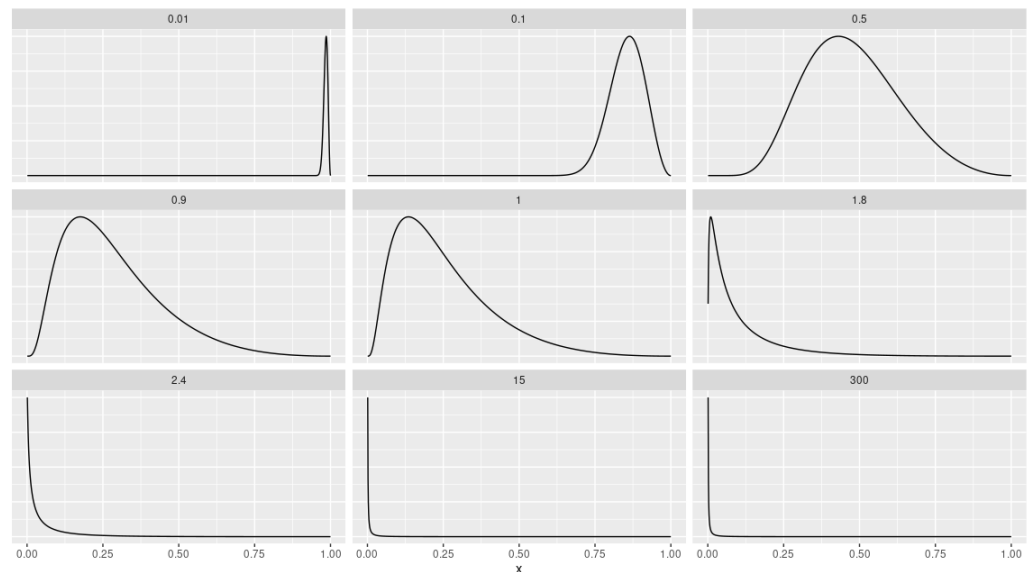


Figure 2. Density function of the AU distribution by varying the parameter α (displayed at the top of each chart). Whereas $B \sim \text{BN}(1) \rightarrow B^2 \sim \chi_3^2$, then the AU model was generated from $X = e^{-\alpha|B|}$.

Proposition 2. If $X \sim AU(\alpha)$, then its CDF is given by

$$F_X(x | \alpha) = 2\Phi\left(\frac{\ln(x)}{\alpha}\right) - 2\left(\frac{\ln(x)}{\alpha}\right)\phi\left(\frac{\ln(x)}{\alpha}\right).$$

Proof. By definition, the CDF is:

$$F_X(x | \alpha) = \int_0^x \frac{2}{t\alpha} \left(\frac{\ln(t)}{\alpha}\right)^2 \phi\left(\frac{\ln(t)}{\alpha}\right) dt. \tag{4}$$

By making the change of variables:

$$\begin{cases} u = \frac{\ln(t)}{\alpha} \Rightarrow e^{u\alpha} = t \\ du = \frac{1}{\alpha t} dt \Rightarrow \alpha e^{u\alpha} du = dt \end{cases},$$

then substituting into Equation (4) and reducing expressions algebraically, we get to:

$$F_X(x | \alpha) = 2 \int_{-\infty}^{\frac{\ln(x)}{\alpha}} u^2 \phi(u) du.$$

By calculating the integral, we find:

$$F_X(x | \alpha) = 2 \left[-u\phi(u) \Big|_{-\infty}^{\ln(x)/\alpha} + \int_{-\infty}^{\ln(x)/\alpha} \phi(u) du \right] \\ = 2 \left[-\left(\frac{\ln(x)}{\alpha}\right)\phi\left(\frac{\ln(x)}{\alpha}\right) + \Phi\left(\frac{\ln(x)}{\alpha}\right) \right].$$

Then, by multiplying and commuting, we get to the expression of Proposition 2. \square

Additionally, if X denotes the monitored variable, then the PDF of X is given by (2). Also, consider that the probability of false alarm (known as type I error) is π . Thus, we get to:

$$\mathbb{P}(X < \text{LCL} | \alpha) = \mathbb{P}(X > \text{UCL} | \alpha) = \pi/2,$$

in which α is the in-control process parameter (that is, the parameter that controls the quality characteristic based on the in-control state), and LCL and UCL are the lower and upper control chart limits, respectively. Given the CDF $F_X(x | \alpha)$, then the quantile function of X is defined by $Q(p | \alpha) = F_X^{-1}(p | \alpha)$, $0 < p < 1$, which can be obtained by setting to zero and solving (numerically) for x the following equation:

$$\Phi\left(\frac{\ln(x)}{\alpha}\right) - \left(\frac{\ln(x)}{\alpha}\right)\phi\left(\frac{\ln(x)}{\alpha}\right) - \frac{p}{2}, \quad \text{for } 0 < p < 1.$$

Following [23], the control limits and centerline (CL) of the proposed control chart for unit data based on the AU distribution or, simply, AU control chart, are given by

$$\text{LCL} = Q(\pi/2 | \alpha), \quad \text{CL} = \mathbb{E}[X | \alpha], \quad \text{UCL} = Q(1 - \pi/2 | \alpha),$$

in which $Q(\cdot)$ is the quantile function of the $AU(\alpha)$ distribution.

Proposition 3. *If $X \sim AU(\alpha)$, then its moment-generating function (MGF) is given by*

$$\psi_X(t | \alpha) = 2 \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{\left(\frac{k^2 \alpha^2}{2}\right)} \left[(1 + k^2 \alpha^2)(1 - \Phi(k\alpha)) - k\alpha\phi(k\alpha) \right].$$

Proof. By definition, the MGF is:

$$\psi_X(t | \alpha) = \mathbb{E}[e^{tx}] = \int_0^1 e^{tx} \frac{2}{x\alpha} \left(\frac{\ln(x)}{\alpha}\right)^2 \phi\left(\frac{\ln(x)}{\alpha}\right) dx. \tag{5}$$

By making the following change of variables:

$$\begin{cases} u = \frac{\ln(x)}{\alpha} & \Rightarrow & e^{u\alpha} = x \\ du = \frac{1}{\alpha x} dx & \Rightarrow & \alpha e^{u\alpha} du = dx \end{cases},$$

then substituting and simplifying into Equation (5), we get to:

$$\psi_X(t | \alpha) = 2 \int_{-\infty}^0 e^{te^{u\alpha}} u^2 \phi(u) du \\ = 2 \int_{-\infty}^0 \sum_{k=0}^{\infty} \frac{t^k e^{u\alpha k}}{k!} u^2 \phi(u) du.$$

Working algebraically, we obtain:

$$\psi_X(t | \alpha) = 2 \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{\left(\frac{\alpha^2 k^2}{2}\right)} \int_{-\infty}^0 u^2 \frac{1}{\sqrt{2\pi}} e^{\left(\frac{-(u-\alpha k)^2}{2}\right)} du.$$

By making the following change of variables: $h = u - \alpha k, dh = du$; then substituting it into the previous equation, we get to:

$$\psi_X(t | \alpha) = 2 \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{\left(\frac{\alpha^2 k^2}{2}\right)} \int_{-\infty}^{-\alpha k} (h + \alpha k)^2 \phi(h) dh.$$

Then, by solving the integral and adjusting algebraically, we get to the expression of Proposition 3. \square

The pseudo-code presented in Algorithm 1 describes the important steps for the generation of random (in fact, pseudo-random) numbers from the $AU(\alpha)$ distribution. Further proofs are attached under Appendix B.

Algorithm 1 Random number generation from the $AU(\alpha)$ model.

Step 1. Generate a random number $x_1 \sim \chi_3^2$.

Step 2. Generate a random number $u \sim \text{Uniform}(0,1)$. If $u \leq 1/2$, set $v = \sqrt{x_1}$; otherwise, $v = -\sqrt{x_1}$.

Step 3. Based on the numbers obtained, generate $y = \alpha|v|$, in which α is a (positive) scale parameter and $|v|$ follows a Bimodal Half-Normal (BHN) distribution.

Step 4. Conclude with the number generated by Step 3 as a negative power of base e , that is, $x = e^{-y} = e^{-\alpha|v|} \in [0, 1]$.

Step 5. Repeat Steps 1–4 n times to obtain a random sample of size n from the $AU(\alpha)$ model.

3. Inference

In this section, the parameter estimation adopting the UMVUE and MLE approaches are discussed. At first, it will be demonstrated that the UMVUE can be obtained straightforwardly, since the proposed AU distribution is part of the exponential family. Later, the MLE will also be discussed, which will help to estimate not only the point estimation of the α parameter, but also the interval estimation. We enrolled the reasoning considering the asymptotic convergence in distribution of the parameter estimator, as well as adapted a transformation that ensures that the interval of the parameter will always be on its domain (the delta method). The delta transformation procedure will enable the correct inferences and the standard error calculation associated with the parameter estimate. Later on, a simulation study to illustrate these theoretical results is presented.

3.1. UMVUE through the Exponential Family

Many of the distributions used in statistics belong to the exponential family, thereby implying in a considerable advantage over other models that do not belong to this family. Such an advantage is significantly declared when it comes to calculating the statistic $T(\mathbf{X})$ of a random sample $\mathbf{X} = (X_1, X_2, \dots, X_n)$. Next, it is shown that the proposed $AU(\alpha)$ distribution belongs to this family.

A random variable X is said to belong to the *one-parameter exponential family* if its associated PDF $f(\cdot | \theta)$ can be written in the form of:

$$f(x | \theta) = \exp\{c(\theta)T(x) + d(\theta) + S(x)\}.$$

Let $X \sim AU(\alpha)$, then the PDF of X can be written in exponential form as it follows:

$$f(x | \alpha) = \exp \left\{ -\frac{1}{2\alpha^2} [\ln(x)]^2 - 3 \ln(\alpha) + \ln \left(\frac{[\ln(x)]^2}{x\sqrt{2\pi}} \right) \right\}.$$

Then, X belongs to the one-parameter exponential family if we define:

$$c(\alpha) = -\frac{1}{2\alpha^2}, \quad T(x) = [\ln(x)]^2, \quad d(\alpha) = -3 \ln(\alpha), \quad S(x) = \ln \left(\frac{[\ln(x)]^2}{x\sqrt{2\pi}} \right).$$

Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ be an observation (or realization) of the random sample $\mathbf{X} = (X_1, X_2, \dots, X_n)$, with $X_i \sim AU(\alpha)$, for $i = 1, 2, \dots, n$. Then, the joint PDF presented in exponential form is

$$f(\mathbf{x} | \alpha) = \exp \left\{ -\frac{1}{2\alpha^2} \sum_{i=1}^n [\ln(x_i)]^2 - 3n \ln(\alpha) + \sum_{i=1}^n \ln \left(\frac{[\ln(x_i)]^2}{x_i\sqrt{2\pi}} \right) \right\},$$

from which it can be concluded that the statistic $T(\mathbf{X}) = \sum_{i=1}^n [\ln(X_i)]^2$ is sufficient and complete, once the AU distribution is part of the exponential family.

Proposition 4. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be a random sample, with $X_i \sim AU(\alpha)$, for $i = 1, 2, \dots, n$, and $T(\mathbf{X}) = \sum_{i=1}^n [\ln(X_i)]^2$. Then,

$$W_n = \frac{1}{\alpha^2} T(\mathbf{X}) \sim \chi_{3n}^2.$$

Proof. If $G = \left[\frac{\ln(X)}{\alpha} \right]^2$, then $G \sim \chi_3^2$. Thus, n independent and identically distributed samples of G will have the sum of $n \chi_3^2$, which will result in a chi-squared distribution with degrees of freedom equal to $3n$, that is, χ_{3n}^2 , since

$$\begin{aligned} F_G(g) &= \mathbb{P}(G \leq g) = \mathbb{P} \left(\left[\frac{\ln(X)}{\alpha} \right]^2 \leq g \right) = \mathbb{P} \left(-\sqrt{g} \leq \frac{\ln(X)}{\alpha} \leq \sqrt{g} \right) \\ &= \mathbb{P}(-\alpha\sqrt{g} \leq \ln(X) \leq \alpha\sqrt{g}) = \mathbb{P}(\ln(X) \leq \alpha\sqrt{g}) - \mathbb{P}(\ln(X) \leq -\alpha\sqrt{g}) \\ &= 1 - \mathbb{P}(\ln(X) \leq -\alpha\sqrt{g}) = 1 - \mathbb{P}(X \leq e^{-\alpha\sqrt{g}}) = 1 - F_X(e^{-\alpha\sqrt{g}}), \end{aligned}$$

so,

$$\begin{aligned} f_G(g) &= \frac{dF_G(g)}{dg} = f_X(e^{-\alpha\sqrt{g}}) (e^{-\alpha\sqrt{g}}) \left(\frac{\alpha}{2\sqrt{g}} \right) \\ &= \frac{2}{\alpha e^{-\alpha\sqrt{g}}} \left(\frac{-\alpha\sqrt{g}}{\alpha} \right)^2 \phi \left(\frac{-\alpha\sqrt{g}}{\alpha} \right) e^{-\alpha\sqrt{g}} \frac{\alpha}{2\sqrt{g}} \\ &= \frac{1}{\sqrt{g}} (\sqrt{g})^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{g})^2}{2}} = \frac{1}{\sqrt{2\pi}} g^{1/2} \exp(-g/2) \equiv \chi_3^2. \end{aligned}$$

□

Proposition 5. Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ be a random sample, with $X_i \sim AU(\alpha)$, for $i = 1, 2, \dots, n$, and $T(\mathbf{X}) = \sum_{i=1}^n [\ln(X_i)]^2$. Then,

$$S(\mathbf{X}) = \frac{\Gamma(\frac{3n}{2})\sqrt{2}}{\Gamma(\frac{3n+1}{2})} \sqrt{T(\mathbf{X})}$$

is an unbiased estimator of α .

Proof. First, remember that if $X \sim \text{Gamma}(a, b)$ distribution, then $\mathbb{E}[X^k] = \frac{\Gamma(a+b)}{b^k \Gamma(a)}$. Since the α parameter is observed to be squared, it will be necessary to apply it to find an unbiased estimator. So, considering the random variable $W_n^{1/2}$ (with W_n as defined in Proposition 4), it follows that:

$$\mathbb{E}[(W_n)^{1/2}] = \frac{\Gamma\left(\frac{3n}{2} + \frac{1}{2}\right)}{2^{1/2} \Gamma\left(\frac{3n}{2}\right)},$$

so,

$$\begin{aligned} \mathbb{E}\left[\left(\frac{1}{\alpha^2} T(\mathbf{X})\right)^{1/2}\right] &= \frac{\Gamma\left(\frac{3n}{2} + \frac{1}{2}\right)}{2^{1/2} \Gamma\left(\frac{3n}{2}\right)} \\ \mathbb{E}\left[\underbrace{\sqrt{T(\mathbf{X})} \frac{\Gamma\left(\frac{3n}{2}\right) \sqrt{2}}{\Gamma\left(\frac{3n}{2} + \frac{1}{2}\right)}}_{S(\mathbf{X})}\right] &= \alpha. \end{aligned}$$

□

Remark 3. Considering the two previous propositions and resorting to the Lehmann-Scheffé theorem, one can conclude that $S(\mathbf{X})$ is UMVUE for α .

3.2. Estimation using the Maximum Likelihood Method

Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ be a realization of the random sample $\mathbf{X} = (X_1, X_2, \dots, X_n)$ taken from the $\text{AU}(\alpha)$ distribution. Then, the log-likelihood function is given by

$$\ell(\alpha) = \text{constant} - 3n \ln(\alpha) - \sum_{i=1}^n \ln(x_i) + 2 \sum_{i=1}^n \ln(\ln(x_i)) - \frac{1}{2\alpha^2} \sum_{i=1}^n [\ln(x_i)]^2.$$

The MLE of α , i.e., $\hat{\alpha}$, is found by solving the following equation:

$$\frac{d\ell(\alpha)}{d\alpha} = -\frac{3n}{\alpha} + \frac{1}{\alpha^3} \sum_{i=1}^n [\ln(x_i)]^2 = 0,$$

resulting

$$\hat{\alpha} = \left\{ \frac{1}{3n} \sum_{i=1}^n [\ln(x_i)]^2 \right\}^{1/2}.$$

On the other hand, the second derivative of $\ell(\alpha)$ evaluated at $\alpha = \hat{\alpha}$ is negative, therefore concluding that $\hat{\alpha}$ is MLE for α .

It is known that, under certain regularity conditions,

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{D} N\left(0, I^{-1}(\alpha)\right),$$

in which $I(\alpha) = -\mathbb{E}\left[\frac{d^2\ell(\alpha)}{d\alpha^2}\right] = \frac{6n}{\alpha^2}$.

A two-sided $100(1 - \pi)\%$ confidence interval for α can be calculated by

$$\left[\hat{\alpha} - z_{1-\pi/2} \sqrt{\text{Var}[\hat{\alpha}]}, \hat{\alpha} + z_{1-\pi/2} \sqrt{\text{Var}[\hat{\alpha}]} \right], \tag{6}$$

in which z_q is the q -th percentile of the standard normal distribution. The variance of $\hat{\alpha}$ can be approximated by the inverse of the observed Fisher information, as

$$\text{Var}[\hat{\alpha}] = I^{-1}(\hat{\alpha}) = \frac{\hat{\alpha}^2}{6n}.$$

Since α is a positive value and we cannot guarantee that the lower limit of the interval (6) is positive, we resort to the delta method to remedy such situation. For this, we define the function $g : [0, \infty) \rightarrow \mathbb{R}$ as $g(\alpha) = \ln(\alpha)$, and knowing that

$$\sqrt{n}(g(\hat{\alpha}) - g(\alpha)) \xrightarrow{D} N\left(0, I^{-1}(\alpha) \left[\frac{dg(\alpha)}{d\alpha}\right]^2\right),$$

we can, then, obtain an approximate two-sided $100(1 - \pi)\%$ confidence interval for α through

$$\left[\hat{\alpha} \exp\left(-\frac{z_{1-\pi/2}}{\sqrt{6n}}\right), \frac{\hat{\alpha}}{\exp\left(-\frac{z_{1-\pi/2}}{\sqrt{6n}}\right)} \right]. \tag{7}$$

3.3. Simulation Study

In order to illustrate the presented inferences for the estimation of the AU distribution, the MLE versus the UMVUE are compared (via simulation study) in this subsection. Moreover, we considered the scenarios in which the parameter $\alpha = \{0.1, 0.3, 0.5, 0.7, 1.1, 1.5\}$, considering sample sizes $n = \{100, 200, 500\}$, through the Monte Carlo method with $N = 1000$ repetitions. This entire procedure took into account the random number generator for the $AU(\alpha)$ distribution shown in Algorithm 1. All analyses carried out in this study adopted the open-source R software [24].

For the performance comparison of the proposed estimators (MLE and UMVUE), since the true parameter value is known, the bias and mean squared error (MSE) metrics were adopted, and they are defined, respectively, as it follows:

$$\text{Bias}(\alpha) = \frac{1}{N} \sum_{i=1}^N (\hat{\alpha}_i - \alpha) \quad \text{and} \quad \text{MSE}(\alpha) = \frac{1}{N} \sum_{i=1}^N (\hat{\alpha}_i - \alpha)^2,$$

in which $\hat{\alpha}_i$ is the estimate for α in the i -th iteration (point estimation). Additionally, based on the asymptotic results presented in this study, we also calculated the 95% confidence interval (CI) length by adopting the delta method from Equation (7) (interval estimation). That is, it analyzed the average of all the upper limits of the 95% confidence interval, as well as the average of all the lower limits, and then calculated their difference.

Table 1 presents the obtained average estimates (AvE) of the α parameter, for each sample size n , as well as the corresponding bias, MSE and 95% CI length (this last one only for MLE) results.

The asymptotic convergence of the MLE towards the robustness is noticed as the sample size increases. In addition, both MLE and UMVUE's bias and MSE are small and tend to decrease as n gets larger. On the other hand, the CI length also decreases as the sample size increases.

Finally, regarding the robustness of the estimators, the difference between the MLE and UMVUE estimates was taken, considering each different sample size n . Then, the interquartile range (IQR) was calculated per sample size group. That is, $\text{IQR}^{(ni)}\left(\hat{\alpha}_{1\text{MLE}}^{(ni)} - \hat{\alpha}_{1\text{UMVUE}}^{(ni)}, \dots, \hat{\alpha}_{j\text{MLE}}^{(ni)} - \hat{\alpha}_{j\text{UMVUE}}^{(ni)}, \dots, \hat{\alpha}_{6\text{MLE}}^{(ni)} - \hat{\alpha}_{6\text{UMVUE}}^{(ni)}\right)$, in which $ni = \{100, 200, 500\}$ and $\alpha_j = \{\alpha_1 = 0.1, \alpha_2 = 0.3, \dots, \alpha_6 = 1.5\}$. For instance, the IQR for $n = 100$ was 0.00053, whereas for $n = 200$ and $n = 500$, it went down to 0.00025 and 0.00012, respectively. This points out, in short, that as the sample size gets larger, the error range gets smaller, regardless of the value of the α parameter.

Table 1. AvE, bias, MSE and 95% CI length (only for MLE) for the proposed estimators (MLE and UMVUE) of the single parameter (α) of the AU distribution, considering different sample sizes (n).

n	α	MLE				UMVUE		
		AvE	Bias	MSE	CI Length	AvE	Bias	MSE
100	0.1	0.0998	−0.0002	1.6930×10^{-5}	0.0160	0.0999	-8.2264×10^{-5}	1.6165×10^{-5}
200		0.0999	-9.8758×10^{-5}	8.7306×10^{-6}	0.0113	0.0999	-5.7124×10^{-5}	8.7314×10^{-6}
500		0.0999	-3.3400×10^{-6}	3.5542×10^{-6}	0.0071	0.1000	1.3327×10^{-5}	3.5555×10^{-6}
100	0.3	0.2996	−0.0004	0.0002	0.0480	0.2999	-8.0656×10^{-5}	0.0002
200		0.2997	−0.0003	7.8575×10^{-5}	0.0339	0.2998	−0.0002	7.8582×10^{-5}
500		0.2999	-1.0020×10^{-5}	3.1987×10^{-5}	0.0214	0.3002	0.0002	3.0979×10^{-5}
100	0.5	0.4994	−0.0006	0.0004	0.0800	0.4999	−0.0001	0.0004
200		0.4997	−0.0003	0.0002	0.0565	0.4997	−0.0003	0.0002
500		0.4999	-1.6700×10^{-5}	8.8855×10^{-5}	0.0357	0.5000	6.6637×10^{-5}	8.8888×10^{-5}
100	0.7	0.6992	−0.0008	0.0008	0.1120	0.6998	−0.0002	0.0008
200		0.6993	−0.0007	0.0004	0.0791	0.6996	−0.0004	0.0004
500		0.6999	-2.3380×10^{-5}	0.0001	0.0501	0.7000	9.3291×10^{-5}	0.0001
100	1.1	1.0987	−0.0013	0.0020	0.1760	1.0997	−0.0003	0.0020
200		1.0989	−0.0011	0.0010	0.1244	1.0994	−0.0006	0.0010
500		1.0999	-3.6741×10^{-5}	0.0004	0.0787	1.1001	0.0001	0.0004
100	1.5	1.4983	−0.0017	0.0038	0.2400	1.4996	−0.0004	0.0038
200		1.4985	−0.0015	0.0019	0.1696	1.4991	−0.0009	0.0019
500		1.4999	-5.0101×10^{-5}	0.0008	0.1073	1.5002	0.0002	0.0007

4. Real-World Exemplifications

In this section, two applications adopting the AU distribution with real-world issues are exemplified. The first case is related to the dynamics of the Chilean inflation in the post-military dictatorship period. The second case pertains to the relative humidity of the air in the northern Chilean city of Copiapó (Atacama region).

The Chilean inflation data are recorded annually, whose values considered the range from 1992 to 2021. These are based on the period after the military dictatorship of 1973–1990. It was analyzed the dynamics of the inflation data (in %), which were standardized by min-max transformation, resulting in a unit response variable (value between zero and one). The years 1990 and 1991 were excluded, since they are considered to be a period of transition. Then, the total amount of observations was of 30 years (from 1992 to 2021).

On the other hand, the relative air humidity data cover the period from February 2015 to October 2022, with a one-hour recording format (104,415 observations). Then, this data set was transformed into daily maximum observation (6226 observations).

4.1. Chilean Inflation (Post-Military Era)

Figure 3 presents the dynamics of the Chilean inflation in the post-military dictatorship period, demonstrating stability between the years of 1999 and 2008. The right panel displays the time series of inflation, in which time is measured in years, from year 1 (1992) to year 30 (2021). The left panel depicts the accumulation of the values throughout the time series, in which a predominant trend is shown around 0.1 of the inflation rate.

Once the empirical dynamics of these data was analyzed, the most common unit distributions, presented in the statistical literature, were fitted. The upper panel of Figure 4 illustrates the histogram for the inflation data, in which it is compared with different fitted densities based on the MLE: AU, beta (BE), Kumaraswamy (KUM), logit-normal (LOGITNO), simplex (SIMPLEX), unit-half-normal (UHN), and unit-Lindley (ULINDLEY). The lower panel of the same figure presents the fitted CDFs superimposed to the empirical CDF (ECDF).

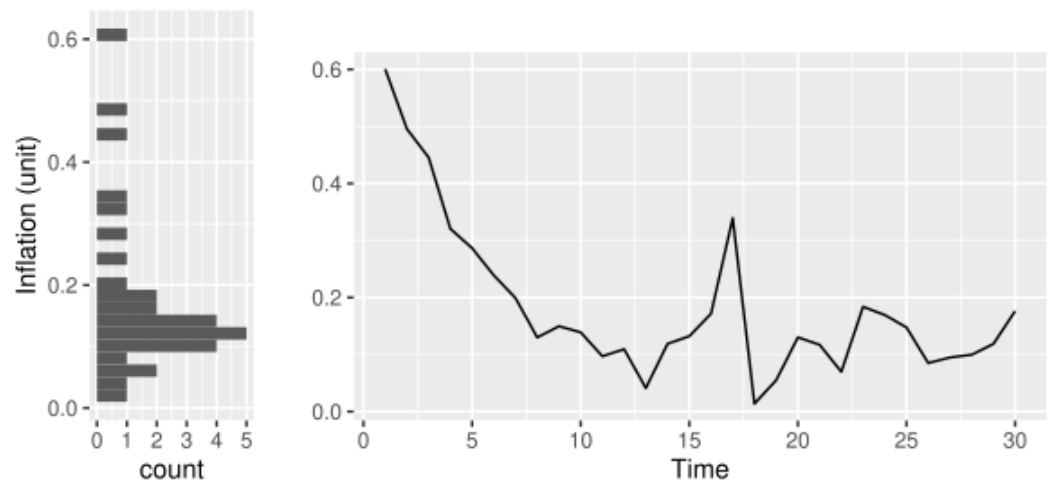


Figure 3. Chilean inflation in the period 1992–2021 (post-military era). The histogram on the left presents a skewness of the data. The dynamics is represented in the right panel, in which a disturbance (outlier) is observed in the year 2008 (observation #17).

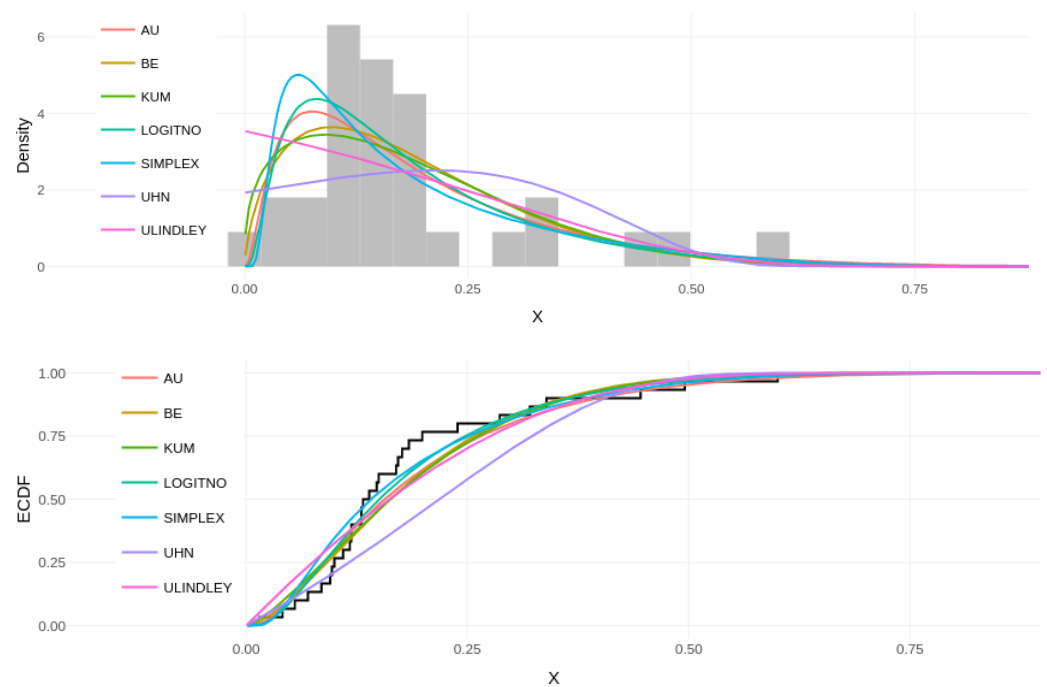


Figure 4. Estimated densities superimposed to the histogram (top-chart), and estimated CDFs superimposed to the ECDF (bottom-chart) (Chilean inflation data).

In order to quantify the performance of the fitted models, the Akaike Information Criterion (AIC) [25] and the Bayesian (or Schwarz) Information Criterion (BIC) [26] were analyzed. The obtained results (see Table 2) indicated the AU model as the best-fitted model to this data set. In addition, it is possible to make an inference about the average of the phenomenon, that is, the expectation of the $AU(\hat{\alpha} = 1.2059)$ model, resulting in $\mathbb{E}[X_{\text{Inflation}}] = 0.1948$. In other words, the average Chilean inflation, in post-military era, is of 19.49%.

In the following subsection, it is illustrated the performance of the AU model when adopting a high-frequency data set originated from the relative humidity from a city located in the Atacama Desert.

Table 2. Parameter estimates, AIC and BIC values (Chilean inflation data). S.E. = standard error.

Model	Parameter Estimate (S.E.)	AIC	BIC
AU(α)	$\hat{\alpha} = 1.205943$ (0.008079)	-47.89	-46.49
BE(μ, σ)	$\hat{\mu} = 0.185857$ (0.000496) $\hat{\sigma} = 0.314688$ (0.001304)	-44.58	-41.78
KUM(μ, σ)	$\hat{\mu} = 1.370127$ (0.045522) $\hat{\sigma} = 7.968427$ (7.750459)	-43.63	-40.83
LOGITNO(μ, σ)	$\hat{\mu} = 0.150323$ (0.000457) $\hat{\sigma} = 0.916938$ (0.014013)	-46.23	-43.43
SIMPLEX(μ, σ)	$\hat{\mu} = 0.182462$ (0.000584) $\hat{\sigma} = 2.854833$ (0.135834)	-43.17	-40.37
UHN(σ)	$\hat{\sigma} = 0.413894$ (0.002855)	-33.62	-32.22
ULINDLEY(μ)	$\hat{\mu} = 0.186834$ (0.000575)	-41.99	-40.58

4.2. Water Monitoring in Air Humidity

The hydrological regime of the main rivers of Atacama is characterized by ice sources: water flows from the peaks following the melting of snowfall, glaciers, and permafrost located in the upper parts of the Andes range. In the context of climate change, it is, therefore, essential to understand the hydrological cycle of these regions, in order to set up a sustainable management policy to them. Understanding the hydrological cycle requires the implementation of tools for forecasting river flows, relative humidity, groundwater reservoirs, or any other water-related quantity monitoring, which inevitably demands an in-depth knowledge with respect to the physical phenomena that rule the entire hydrological cycle and, more precisely, the complex interaction between atmosphere, climate, landforms, ice, snow and river flows.

Additionally, a unique phenomenon called *Camanchaca* happens, which consists in a fog passing by the Copiapó city, recurrent only between midnight to around 10 a.m. Here, we demonstrate the variation of the relative humidity of Copiapó city, proposing a methodology that can be efficient, adjustable to these data. Using the daily maximum relative humidity, six different unit distributions were compared: AU, BE, KUM, LOGITNO, SIMPLEX, and UHN, as shown in Figure 5.

After comparing the commonly used unit models, we demonstrate the advantage of fitting the AU model over others (visually). Table 3 confirms the best fit of the AU model, based on information criteria (AIC and BIC), as well as depicts the estimation of the parameter(s) of each model.

Table 3. Parameter estimates, AIC and BIC values (relative air humidity data).

Model	Parameter Estimate (S.E.)	AIC	BIC
AU(α)	$\hat{\alpha} = 0.1092$ (3.1902×10^{-7})	-14,023.49	-14,016.76
BE(μ, σ)	$\hat{\mu} = 0.8476$ (1.2027×10^{-6}) $\hat{\sigma} = 0.2410$ (4.1119×10^{-6})	-13,927.89	-13,914.41
KUM(μ, σ)	$\hat{\mu} = 9.4004$ (0.0141) $\hat{\sigma} = 2.3882$ (0.0019)	-13,605.90	-13,592.43
LOGITNO(μ, σ)	$\hat{\mu} = 0.8693$ (3.1376×10^{-6}) $\hat{\sigma} = 1.2299$ (1.2148×10^{-4})	-7600.43	-7586.95
SIMPLEX(μ, σ)	$\hat{\mu} = 0.9735$ (1.2959×10^{-6}) $\hat{\sigma} = 94.0480$ (0.7103)	32,477.13	32,490.61
UHN(σ)	$\hat{\sigma} = 99.9900$ (6.5334×10^{-7})	5,101,018,733.13	5,101,018,739.86

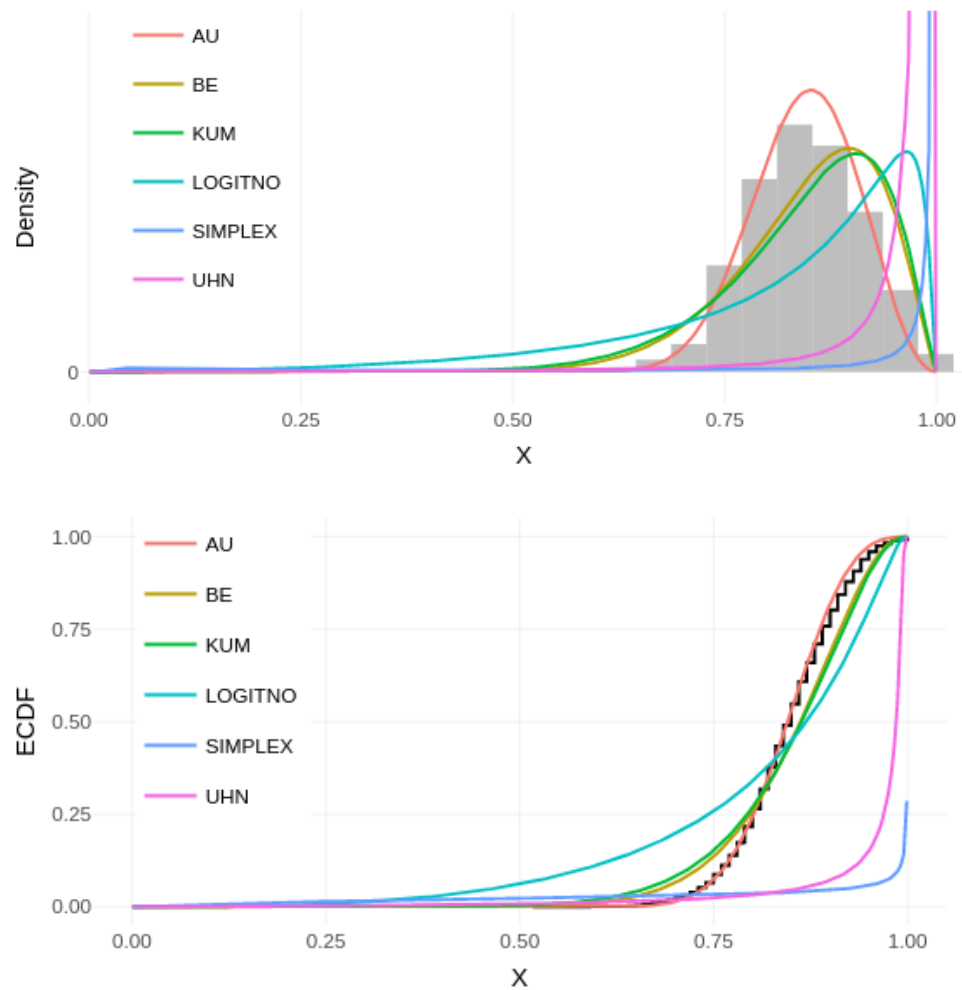


Figure 5. Estimated densities superimposed to the histogram (**top-chart**), and estimated CDFs superimposed to the ECDF (**bottom-chart**) (relative air humidity data).

After obtaining the parameter estimate for α , the AU model (best-fitted model) was used to construct a Statistical Process Control (SPC) chart [27], by calculating a tolerance upper-lower bound. Moreover, the Highest Density Interval (HDI) was adopted, considering a confidence degree of 99%, to monitor the daily maximum relative humidity records (as displayed by Figure 6).

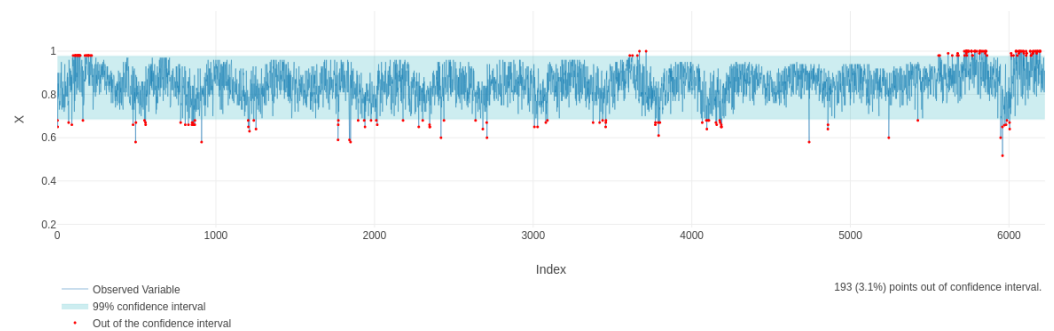


Figure 6. SPC control chart, considering a 99% of tolerance based on the AU model fitted to the daily maximum relative humidity of Copiapó city, Chile, from 1 February 2015 to 4 October 2022. It is observed that 193 days (3.1%) presented anomaly values (out-of-control signals). The obtained control limits were: LCL = 68.56% and UCL = 97.73%.

The expected daily maximum water relative humidity is of 84.23% (based on the fitted AU model). The obtained control limits, considering a confidence (or tolerance) level of 99%, were: LCL = 68.56% and UCL = 97.73%. Thus, the control chart based on the AU model (AU control chart) is another exciting and valuable alternative to some well-known SPC tools, which enlightens the forecasting and opens new doors to discuss extreme events in the Atacama water particles monitored by probabilistic reasoning.

5. Conclusions

This study showed the competitiveness of the developed Theorem 1 (Equation (1)), which enables for a great class of distributions that belong all to the exponential family. As an exemplification, we adopted the special case for $k = 1$, which is equivalent to the moment of order two of the standard normal distribution, and after some transformations, we developed the Alpha-Unit (AU) distribution. Also, we dedicated to the unit range, given the importance of this stochasticity representation.

Unit distributions are useful for values that oscillate between zero and one, such as fractions, proportions and rates, among others, or for a set of values in which there is a minimum or maximum limitation, resorting to standardization through the min-max transformation. Most distributions of this type come from transforming a random variable with certain distribution so that it takes values between zero and one, as in the case of unit-Lindley distribution [9], which comes from the Lindley distribution [28,29].

There are numerous studies based on (e.g., unit) distributions, by extending a model and applying it to several areas [11,14,16]. In this study, we introduced and showed the competitiveness of the AU distribution, especially for data with a range greater than 0.4, or which present high asymmetry and low decay. Further studies shall investigate this hypothesis in a wider amount of data sets (through different sorts of wide data range). Additionally, an implementation of this model adopting hierarchical estimation and spatio-temporal dependence would be useful for forecast/predictable problems.

Author Contributions: Conceptualization and methodology, D.E.-O. and D.C.d.N.; software, M.S.C.-A. and L.B.-B.; validation, formal analysis, investigation, resources, data curation, M.S.C.-A. and D.C.d.N.; writing—original draft preparation, D.E.-O., D.C.d.N., M.S.C.-A. and L.B.-B.; writing—review and editing, P.H.F.d.S.; visualization, D.C.d.N.; supervision, project administration, funding acquisition, D.E.-O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universidad de Atacama grant number ATA1956–CC88433.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All adopted data and R script developed in this study are available at <https://github.com/ProfNascimento/AlphaUnit> (accessed on 12 November 2022).

Acknowledgments: This study was partially supported by the Vicerrectoría de Investigación y Postgrado (VRIP) and Dirección de Postgrado of the Universidad de Atacama (UDA). The author David Elal-Olivero was supported by the DIUDA REGULAR project No. 22409 from the Universidad de Atacama, Chile.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

This appendix shows the proof that for a random variable

$$X \sim \text{BN}(k) \quad \rightarrow \quad X^2 \sim \chi_{2k+1}^2.$$

Then,

$$F_{X^2}(x) = \mathbb{P}(X^2 \leq x) = \mathbb{P}(-\sqrt{x} \leq X \leq \sqrt{x}) = 2\mathbb{P}(X \leq \sqrt{x}) - 1 = 2F_X(\sqrt{x}) - 1.$$

It follows that

$$f_{X^2}(x) = 2f_X(\sqrt{x}) \frac{1}{2\sqrt{x}} = \frac{1}{c} (\sqrt{x})^{2k} \phi(\sqrt{x}) \frac{1}{\sqrt{x}} = \frac{1}{\prod_{j=1}^k (2j-1)} (\sqrt{x})^{2k-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}}.$$

Knowing that $\Gamma\left(\frac{2k+1}{2}\right) = \prod_{j=1}^k (2j-1) \frac{\sqrt{\pi}}{2^k}$, then

$$\begin{aligned} f_{X^2}(x) &= \frac{1}{\prod_{j=1}^k (2j-1)} x^{\frac{2k-1}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} = \frac{\sqrt{\pi}}{2^k \Gamma\left(\frac{2k+1}{2}\right)} x^{\frac{2k-1}{2}} e^{-\frac{x}{2}} \\ &= \frac{1}{\Gamma\left(\frac{2k+1}{2}\right)} 2^{\frac{2k+1}{2}} x^{\frac{2k-1}{2}} e^{-\frac{x}{2}}. \end{aligned}$$

Therefore, $X^2 \sim \chi_{2k+1}^2$.

Besides that, complementation can be taken into account by saying that, considering $W_2 \sim \chi_{2k+1}^2$ and $\mathbb{P}(W_1 = \pm 1) = 1/2$, then $B = W_1 \sqrt{W_2} \sim \text{BN}(k)$.

Let $b \geq 0$, then

$$\begin{aligned} F_B(b) &= \mathbb{P}(B \leq b) = \mathbb{P}(W_1 \sqrt{W_2} \leq b) \\ &= \mathbb{P}(W_1 \sqrt{W_2} \leq b \mid W_1 = 1) \mathbb{P}(W_1 = 1) + \mathbb{P}(W_1 \sqrt{W_2} \leq b \mid W_1 = -1) \mathbb{P}(W_1 = -1) \\ &\stackrel{\text{ind.}}{=} \mathbb{P}((1) \sqrt{W_2} \leq b) \frac{1}{2} + \mathbb{P}((-1) \sqrt{W_2} \leq b) \frac{1}{2}. \end{aligned}$$

Since $b \geq 0$, then $\mathbb{P}((-1) \sqrt{W_2} \leq b) = 1$:

$$\begin{aligned} &= \mathbb{P}(\sqrt{W_2} \leq b) \frac{1}{2} + \frac{1}{2} = \mathbb{P}(|W_2| \leq b^2) \frac{1}{2} + \frac{1}{2} = \mathbb{P}(-b^2 \leq W_2 \leq b^2) \frac{1}{2} + \frac{1}{2} \\ &= \frac{1}{2} \left[\mathbb{P}(X \leq b^2) - \underbrace{\mathbb{P}(X \leq -b^2)}_0 \right] + \frac{1}{2} = \frac{1}{2} \mathbb{P}(X \leq b^2) + \frac{1}{2} = \frac{1}{2} F_X(b^2) + \frac{1}{2} \end{aligned}$$

Therefore,

$$\begin{aligned} f_B(b) &= \frac{dF_B(b)}{db} = \frac{1}{2} f_X(b^2) 2b = b f_X(b^2) = b \frac{1}{\Gamma\left(\frac{2k+1}{2}\right) 2^{\frac{2k+1}{2}}} (b^2)^{\frac{2k+1}{2}-1} e^{-\frac{b^2}{2}} \\ &= b \frac{1}{\Gamma\left(\frac{2k+1}{2}\right) 2^{\frac{2k+1}{2}}} b^{2k-1} e^{-\frac{b^2}{2}} = \frac{1}{\frac{\sqrt{\pi} \prod_{j=1}^k (2j-1)}{2^k} 2^{\frac{2k+1}{2}}} b^{2k} e^{-\frac{b^2}{2}} \\ &= \frac{1}{\prod_{j=1}^k (2j-1)} \frac{b^{2k}}{\sqrt{2\pi}} e^{-\frac{b^2}{2}} = \frac{1}{\underbrace{\prod_{j=1}^k (2j-1)}_c} b^{2k} \phi(b). \end{aligned}$$

Analogously, it is proved for $b < 0$.

Appendix B

The proposed theorem (Theorem 1) will be illustrated considering $k = 1$, to show the origin of the random numbers that generate the AU distribution.

Proposition A1. *If $X \sim \text{BN}(1)$, then $f_X(x) = x^2 \phi(x)$ is a bimodal density function.*

Proof. If $f_X(x)$ is bimodal, it would have two maxima, to which the first and second derivative criteria would be applied:

$$\frac{df_X(x)}{dx} = 0 \Rightarrow$$

$$\frac{d(x^2\phi(x))}{dx} = 2x\phi(x) + x^2[-x\phi(x)] = 2x\phi(x) - x^3\phi(x) = x\phi(x)(2 - x^2) = 0.$$

Then, it can be seen that the solutions for the previous equation would be: $x_1 = 0$, $x_2 = \sqrt{2}$, $x_3 = -\sqrt{2}$. Hence, by applying the second derivative criterion:

$$\frac{d^2f_X(x)}{dx^2} < 0 \Rightarrow$$

$$\frac{d(x\phi(x)(2 - x^2))}{dx} = \phi(x)(2 - x^2) + x[-x\phi(x)](2 - x^2) + x\phi(x)(-2x).$$

Reducing algebraically, we get to:

$$\frac{d^2f_X(x)}{dx^2} = \phi(x)(2 - 5x^2 + x^4) < 0.$$

The only solutions that satisfy the previous inequality are: $x_2 = \sqrt{2}$, $x_3 = -\sqrt{2}$. Therefore, there are two maxima and the BN distribution is bimodal.

□

Definition A1 (Bimodal Half-Normal distribution). Let $Y \sim BN(1)$. If $Q = \alpha|Y|$, with $\alpha > 0$, then we say that Q is distributed according to a Bimodal Half-Normal (BHN) distribution with parameter α , and we denote it by $Q \sim BHN(\alpha)$.

Proposition A2. If $Q \sim BHN(\alpha)$, then the PDF of Q is given by

$$f_Q(q | \alpha) = \frac{2}{\alpha} \left(\frac{q}{\alpha}\right)^2 \phi\left(\frac{q}{\alpha}\right), \quad q > 0.$$

Proof. Since $Q = \alpha|Y|$, with $Y \sim BN(1)$, then

$$F_Q(q) = \mathbb{P}(Q \leq q) = \mathbb{P}(\alpha|Y| \leq q) = \mathbb{P}\left(-\frac{q}{\alpha} \leq Y \leq \frac{q}{\alpha}\right) = 2\mathbb{P}\left(Y \leq \frac{q}{\alpha}\right) - 1 = 2F_Y\left(\frac{q}{\alpha}\right) - 1.$$

Hence, by deriving the previous expression, one has that

$$f_Q(q) = 2f_Y\left(\frac{q}{\alpha}\right) \frac{1}{\alpha} = \frac{2}{\alpha} \left(\frac{q}{\alpha}\right)^2 \phi\left(\frac{q}{\alpha}\right).$$

□

Proposition A3. If $Q \sim BHN(\alpha)$, then

$$X = e^{-Q} \sim AU(\alpha).$$

Proof. Let $X = e^{-Q}$, $0 < x \leq 1$, then

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}\left(e^{-Q} \leq x\right) = \mathbb{P}(-Q \leq \ln(x)) = \mathbb{P}(Q \geq -\ln(x)) \\ &= 1 - \mathbb{P}(Q \leq -\ln(x)) = 1 - F_Q(-\ln(x)). \end{aligned}$$

By deriving the previous expression, we have:

$$f_X(x) = f_Q(-\ln(x)) \frac{1}{x} = \frac{2}{\alpha} \left(\frac{-\ln(x)}{\alpha} \right)^2 \phi \left(\frac{-\ln(x)}{\alpha} \right) \frac{1}{x} = \frac{2}{\alpha x} \left(\frac{\ln(x)}{\alpha} \right)^2 \phi \left(\frac{\ln(x)}{\alpha} \right).$$

□

References

- Fonseca, A.; Ferreira, P.H.; Nascimento, D.C.d.; Fiaccone, R.; Ulloa-Correa, C.; García-Piña, A.; Louzada, F. Water particles monitoring in the atacama desert: SPC approach based on proportional data. *Axioms* **2021**, *10*, 154. [\[CrossRef\]](#)
- Bayer, F.M.; Cintra, R.J.; Cribari-Neto, F. Beta seasonal autoregressive moving average models. *J. Stat. Comput. Simul.* **2018**, *88*, 2961–2981. [\[CrossRef\]](#)
- Ferrari, S.; Cribari-Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **2004**, *31*, 799–815. [\[CrossRef\]](#)
- Kumaraswamy, P. A generalized probability density function for double-bounded random processes. *J. Hydrol.* **1980**, *46*, 79–88. [\[CrossRef\]](#)
- Tadikamalla, P.R.; Johnson, N.L. Systems of frequency curves generated by transformations of logistic variables. *Biometrika* **1982**, *69*, 461–465. [\[CrossRef\]](#)
- Barndorff-Nielsen, O.E.; Jørgensen, B. Some parametric models on the simplex. *J. Multivar. Anal.* **1991**, *39*, 106–116. [\[CrossRef\]](#)
- Mazucheli, J.; Menezes, A.F.B.; Ghitany, M.E. The unit-Weibull distribution and associated inference. *J. Appl. Probab. Stat.* **2018**, *13*, 1–22.
- Mazucheli, J.; Menezes, A.; Fernandes, L.; De Oliveira, R.; Ghitany, M. The unit-Weibull distribution as an alternative to the Kumaraswamy distribution for the modeling of quantiles conditional on covariates. *J. Appl. Stat.* **2020**, *47*, 954–974. [\[CrossRef\]](#)
- Mazucheli, J.; Menezes, A.F.B.; Chakraborty, S. On the one parameter unit-Lindley distribution and its associated regression model for proportion data. *J. Appl. Stat.* **2019**, *46*, 700–714. [\[CrossRef\]](#)
- Bakouch, H.S.; Nik, A.S.; Asgharzadeh, A.; Salinas, H.S. A flexible probability model for proportion data: Unit-half-normal distribution. *Commun. Stat.-Case Stud. Data Anal. Appl.* **2021**, *7*, 271–288. [\[CrossRef\]](#)
- Korkmaz, M.Ç.; Korkmaz, Z.S. The unit log-log distribution: A new unit distribution with alternative quantile regression modeling and educational measurements applications. *J. Appl. Stat.* **2021**, *1*, 1–20. [\[CrossRef\]](#)
- Sagrillo, M.; Guerra, R.R.; Bayer, F.M. Modified Kumaraswamy distributions for double bounded hydro-environmental data. *J. Hydrol.* **2021**, *603*, 127021. [\[CrossRef\]](#)
- Krishna, A.; Maya, R.; Chesneau, C.; Irshad, M.R. The Unit Teissier Distribution and Its Applications. *Math. Comput. Appl.* **2022**, *27*, 12. [\[CrossRef\]](#)
- Guerra, R.R.; Peña-Ramírez, F.A.; Bourguignon, M. The unit extended Weibull families of distributions and its applications. *J. Appl. Stat.* **2021**, *48*, 3174–3192. [\[CrossRef\]](#)
- Aitchison, J.; Brown, J.A.C. *The Lognormal Distribution with Special Reference to Its Uses in Economics*; Cambridge University Press: Cambridge, UK, 1957.
- Korkmaz, M.Ç.; Chesneau, C.; Korkmaz, Z.S. The Unit Folded Normal Distribution: A New Unit Probability Distribution with the Estimation Procedures, Quantile Regression Modeling and Educational Attainment Applications. *J. Reliab. Stat. Stud.* **2022**, *15*, 261–298. [\[CrossRef\]](#)
- Afify, A.Z.; Nassar, M.; Kumar, D.; Cordeiro, G.M. A new unit distribution: properties, inference, and applications. *Electron. J. Appl. Stat. Anal.* **2022**, *15*, 460–484.
- Korkmaz, M.Ç.; Altun, E.; Chesneau, C.; Yousof, H.M. On the unit-Chen distribution with associated quantile regression and applications. *Math. Slovaca* **2022**, *72*, 765–786. [\[CrossRef\]](#)
- Santana-e Silva, J.J.; Cribari-Neto, F.; Vasconcellos, K.L. Beta distribution misspecification tests with application to Covid-19 mortality rates in the United States. *PLoS ONE* **2022**, *17*, e0274781. [\[CrossRef\]](#)
- Stahl, S. The evolution of the normal distribution. *Math. Mag.* **2006**, *79*, 96–113. [\[CrossRef\]](#)
- Limpert, E.; Stahel, W.A. Problems with using the normal distribution—and ways to improve quality and efficiency of data analysis. *PLoS ONE* **2011**, *6*, e21403. [\[CrossRef\]](#)
- Elal-Olivero, D. Alpha-skew-normal distribution. *Proyecciones (Antofagasta)* **2010**, *29*, 224–240. [\[CrossRef\]](#)
- Bayer, F.M.; Tondolo, C.M.; Müller, F.M. Beta regression control chart for monitoring fractions and proportions. *Comput. Ind. Eng.* **2018**, *119*, 416–426. [\[CrossRef\]](#)
- R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
- Akaike, H. On entropy maximization principle. *Appl. Stat.* **1977**, *1*, 27–41.
- Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* **1978**, *6*, 461–464. [\[CrossRef\]](#)
- Montgomery, D.C. *Introduction to Statistical Quality Control*, 6th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2009.
- Lindley, D.V. Fiducial distributions and Bayes' theorem. *J. R. Stat. Soc. Ser. B (Methodol.)* **1958**, *20*, 102–107. [\[CrossRef\]](#)
- Lindley, D.V. *Introduction to Probability and Statistics from a Bayesian Viewpoint, Part II: Inference*; Cambridge University Press: Cambridge, UK, 1965.